ROT Quantum Layer Series — Official Concept Publication (Vol. 1-6) 3편

Author & Original Proposer: Noh Taeyoon (노태윤)
First public release of the ROT Quantum Layer paradigm and its six core components:
ΔΦ (Phase Difference) 기반 변화 구조, 정렬 비용(Alignment Cost),
정합성 기반 AI(Consistency-Driven AI), 신뢰 스코어(Trust Score),
디지털 실존 토큰(DET), ROT Quantum Layer 총론.

신뢰 스코어(Trust Score) 개념의 통합적 정의:
도메인을 초월한 판단·검증·정렬 구조의 새로운 기준

2025

초록(Abstract)

본 논문은 다양한 분야—AI, 데이터 검증, 사회 시스템, 인간 의사결정, 정책 판단 등—에서 공통적으로 요구되는 “신뢰(Trust)”의 기준을 **정량화된 구조(Trust Score)**로 재정의한다.
현대 사회는 데이터의 폭증, 정보의 조작 가능성 증가, AI 판단의 불안정성 등으로 인해 기존의 신뢰 체계가 붕괴되고 있다. 본 연구는 신뢰를 단순한 감정적 개념이 아닌 **정합성(consistency), 출처(origin), 실존(authenticity), 패턴 안정성(stability)**을 기반으로 계산되는 “구조적 속성”으로 재해석한다.
이 논문은 신뢰 스코어를 **도메인 불문, 시스템 중립적 평가값**으로 정의하며, 향후 AI 판단, 디지털 인증, 정책 평가, 금융·보안 시스템 등 모든 영역에서 사용할 수 있는 공통 언어로 제안한다.

---

1. 서론(Introduction)

“신뢰”는 인간·사회·AI·시스템이 작동하기 위한 최소 조건이다.
그러나 현대의 문제는 다음과 같다.

1) 정보의 신뢰 상실

- 온라인 정보의 출처 불명

조작·가짜뉴스의 범람
데이터 왜곡·편향 증가

2) AI의 신뢰 부족

hallucination
출처 미제공
판단 근거 불명
반복 질문 시 답변 불일치

3) 사회·경제적 신뢰 붕괴

올바른 정책 판단 불가
금융·보험 사기 증가
인증 시스템의 취약성
기존 시스템은 **"신뢰"를 감정적·정성적 개념**으로만 다루어
정량적 측정이 불가능했다.
본 연구는 신뢰를 **정량적·구조적·패턴 기반 값**으로 재정의한다.
이것이 본 논문의 핵심이다.

---

2. 신뢰(Trust)의 구조적 재정의

신뢰는 단순히 "믿을 수 있다"는 감각이 아니라
다음 네 요소의 정합적 결과물이다.

2.1 출처 기반 신뢰 (Origin Trust)

·

정보 또는 행동의 실존 여부
실소유자·실행자의 명확성
위조·변조 여부
DET 서명 가능성 여부

2.2 사실 기반 신뢰 (Factual Trust)

사실 여부
검증 가능성
여러 출처 간의 교차 정합성
위조 불가능한 과거 기록 유무

2.3 맥락 기반 신뢰 (Contextual Trust)

상황적 정합성
문맥 흐름의 일관성
주체·역할의 안정성
동일 맥락에서 동일 패턴 유지

2.4 행동 기반 신뢰 (Behavioral Trust)

과거 행동 패턴의 지속성
예측 가능성
ΔΦ 변화 패턴의 안정성
ROT 정렬 비용과의 연결성

본 논문은 위 네 요소를 통합하여
신뢰를 단일한 수치값으로 구성하는 **Trust Score**를 제안한다.

3. 신뢰 스코어(Trust Score)의 개념적 정의

신뢰 스코어는 다음과 같이 정의된다.
"한 정보·행동·데이터가
출처·사실·맥락·행동 측면에서
얼마나 일관된 패턴을 유지하는지를
정량화한 정합성 기반 지표."

이 점수는 도메인에 상관없이 작동한다.
AI 판단
인간 행동 분석
데이터 인증
금융·거래 시스템
뉴스·정보 검증
자율주행·군사 행동 분석
행정·정책 평가
모든 곳에서 동일한 구조로 적용할 수 있다.

4. 신뢰 스코어의 네 축(4-Pillar Trust Model)

4.1 Origin Pillar — 출처 신뢰

정보가 "누구에게서 왔는가"는 신뢰의 1차 조건이다.
출처가 명확하지 않으면
어떤 정보도 신뢰될 수 없다.

4.2 Fact Pillar — 사실 신뢰

사실 여부는 신뢰의 핵심이다.
이는 단순 참·거짓이 아니라
검증 가능성 + 출처 교차 정합성까지 포함한다.

4.3 Context Pillar — 맥락 신뢰

같은 사실도 맥락이 맞지 않으면 신뢰될 수 없다.
따라서 신뢰는 논리적 흐름·문맥적 자연스러움을 포함한다.

4.4 Behavior Pillar — 행동 신뢰

사람·사건·시스템·데이터는
행동 패턴을 통해 가장 정확히 드러난다.
ΔΦ가 안정적이라면
행동 신뢰는 높아진다.
ROT와 직접 연결되는 지점이다.

5. 신뢰 스코어의 구조적 특징

5.1 도메인 불문(Domain-Neutral)

인간, 시스템, AI, 데이터, 정책 등
어느 영역에서도 동일 기준으로 평가할 수 있다.

5.2 정합성 기반(Consistency-Centric)

확률, 감정, 평판 기반이 아니다.
오직 **정합성(Consistency)**으로만 판단한다.

5.3 DET·ROT와 통합 가능

-

출처는 DET
변화 패턴은 ΔΦ
선택 방향은 정렬 비용
이 세 가지가 결합하면
세계 최초의 "정렬 기반 신뢰 AI"가 완성된다.

5.4 표준화 가능(Scalable Standard)

향후 국가·기업·국제기구에서 사용할 수 있는
“통합 신뢰 평가 프레임워크”가 된다.

---

6. 기존 신뢰 평가와의 비교

항목	기존 방식	본 논문에서 정의한 Trust Score
기준	감정, 명성, 경험	출처·사실·맥락·행동 정합성
수치화	어려움	가능
범위	금융 중심	도메인 불문
조작 가능성	높음	낮음(DET 기반)
AI 적합성	낮음	매우 높음
설명 가능성	낮음	정합성 구조로 높은 설명성 제공

Trust Score는 기존 모든 신뢰 평가 모델을
근본적으로 대체 가능한 구조다.

---

7. 본 연구의 기여(Contribution)

✔ 1) 신뢰를 세계 최초로 구조화·정량화

신뢰를 감정이 아닌
“정렬 기반 구조적 속성”으로 정의했다.

✔ 2) 도메인 불문 통합 체계 제안

인간·AI·데이터·정책에 공통 사용 가능한
세계 최초의 신뢰 평가 프레임워크.

✔ 3) DET·ΔΦ·정합성 모델과의 통합

이 논문은 ROT 기반 신뢰 AI의
핵심 중추적 이론이다.

✔ 4) AI 시대의 표준이 될 수 있는 구조

정책·의료·금융·군사·행정 등
모든 분야에서 호출 가능한 단일 기준을 제시한다.

---

8. 결론(Conclusion)

Trust Score는 단순한 점수 체계가 아니라
향후 모든 판단 구조의 공통 기반이 될 수 있는
철학적·구조적 원리이다.
본 논문은 Trust Score를:
출처
사실
맥락
행동
네 요소의 정합성 기반 구조로 정의하며,
AI 시대의 신뢰 문제를 해결하기 위한
새로운 판단 패러다임을 제시한다.
이 구조는 향후 정합성 AI, DET 인증, ROT 정렬 모델과 결합해
완전한 신뢰 기반 AI 생태계를 구성하게 된다.

부록: 공개 범위 안내

※ 본 문서는 신뢰 스코어의 개념·철학적 정의만 포함한다.
※ 실제 점수 계산식, 요인별 가중치, 알고리즘 등
사업 핵심 요소는 고의적으로 공개하지 않는다.
→ 이 논문은 **우선 제안자(창시자)** 기록용 아카이브 버전이다.